NATURAL HISTORY OF COAGULOPATHY IN COVID-19 PATIENTS AND PERSONS VACCINATED AGAINST SARS-COV- DURING THE OMICRON PERIOD.

PASS information

Title	Natural history of coagulopathy in COVID-19 patients and persons vaccinated against SARS-CoV-2 during the Omicron period.
Protocol version identifier	1.0
Date of last version of protocol	12/05/2023
EU PAS register number	
Project aim	The aim of the project is to contextualise the risk of venous and arterial thromboembolic events associated with COVID-19, during the Omicron period and in light of prior SARS-CoV-2 vaccination and infection.
Research objectives	1) To estimate the background incidence rate of venous and arterial thromboembolic events among the general pre-pandemic population. 2) To estimate the incidence rate of venous and arterial thromboembolic events among patients with COVID-19 within 30-, 60-, and 90- and 180-days during the Omicron period, stratified by prior SARS-CoV-2 vaccination and prior infection status. 3) To estimate the incidence rate of venous and arterial thromboembolic events among patients with SARS-CoV-2 vaccination within 30-, 60-, 90- and 180-days, stratified by prior infection status. 4) To estimate the association between clinical risk factors and prior SARS-CoV-2 vaccination on the incidence rate of venous and arterial and impact of thromboembolic events on worsening severity of COVID-19 during the Omicron period.

	5) To estimate incidence rate ratios for venous and arterial thromboembolic events among patients with COVID-19 and different SARS-CoV-2 vaccine doses compared to the background population, using incidence rates estimated in objectives 1 to 3.
Country(-ies) of study	Denmark, Croatia, France, Finland, EMA, DARWIN EU network
Author	Daniel Morales and Denise Umuhire, on behalf of the EHDS2 pilot use case consortium

2 Contents

3 List of abbreviations

Abbreviation	Name
ATC	Anatomical Therapeutic Chemical Classification
CDM	Common Data Model
COVID-19	Coronavirus disease-2019
DVT	Deep vein thrombosis
ECMO	Extracorporeal membrane oxygenation
EHR	Electronic Health Record
ENCePP	European Network of Centres for Pharmacoepidemiology and Pharmacovigilance
ICU	Intensive care unit
LPD	Longitudinal Patient Data
MACE	Major cardiovascular events
OMOP	Observational Medical Outcomes Partnership
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
VTE	Venous thromboembolic events

4 Responsible parties

Croatian Institute of Public Health (HR)	Jakov Vuković Emanuel Brađašević Pero Ivanko
Danish Health Data Authority (DK)	Jesper Kjær Christian Fynbo Christiansen
DARWIN EU Coordination Centre	Marti Catala-Sabate Xintong Li Daniel Prieto Alhambra
European Medicines Agency*	Daniel R. Morales Denise Umuhire
Finnish Insittute of Health and Welfare (FI)	Persephone Doupi Haaramo Peija Jokinen Jukka
Health Data Hub (FR)	Emmanuel Bacry Mario Jendrossek Camille Cloitre Lorien Benda
Health Data Lab (HDL) / Federal Institute for Drugs and Medical Devices (DE)	Katharina Schneider

*EHDS2 Pilot use case lead organisation

5 Amendments and updates

There have been no formal amendments to the protocol so far.

6 Milestones

Milestone	**Date**
Approval of Study Protocol	12/05/2023
Start of data collection	12/05/2023
End of data collection	
Draft report	
Final study report	

7 Rationale and background

7.1 European Health Data Space

HealthData@EU is the European Health Data Space (EHDS) Pilot project that aims to investigate and establish an infrastructure and data ecosystem for the secondary use of health data for research, innovation and better policy making. The HealthData@EU pilot project will assess the ability to scale towards a Union-wide infrastructure as a core component of the European Health Data Space. The proposed study is one of the five use cases selected to test and inform HealthData@EU frameworks.

7.2 Occurrence of venous and arterial thromboembolic events in COVID-19

Coronavirus disease-2019 (COVID-19) may cause both venous and arterial thromboembolic events due to excessive inflammation, platelet activation, endothelial dysfunction, and stasis.[1] A number of observational studies and case series have reported high rates of venous and arterial thromboembolic events among patients hospitalised with COVID-19. In a case series of COVID-19 patients admitted to ICU in the Netherlands, the incidence of thrombotic complications was found to be 31%,[2] while a similar case series from a hospital in Italy found the incidence of thromboembolic events to be 28%.[3] Meanwhile, the rate of venous thromboembolism was found to be as high as 69% for a case series from two French intensive care units (ICU).[4]

Studies assessing the incidence of thromboembolic events in COVID-19 during the early period of the pandemic were conducted in relatively small study populations, partly constrained by data availability at the time, and predominantly focused on hospitalised populations. Consequently, uncertainty remained around the incidence of thromboembolic events among patients with COVID-19. Several larger studies examining coagulopathy risk have since been conducted using routinely collected health data. One study from Sweden demonstrated an increased risk of deep vein thrombosis (DVT) up to 70 days post-COVID-19 diagnosis and an increased risk of pulmonary embolism up to 110 days post-COVID-19 diagnosis.[5] Meanwhile, another study from England reported an increased risk of venous thromboembolism up to 49 weeks post-COVID-19.[6] Studies have also observed that the elevated risk of thromboembolic events associated with COVID-19 infection may be attenuated following SARS-CoV-2 vaccination.[7] SARS-CoV-2 variants have changed over time, with Omicron now widely established as the dominant SARS-CoV-2 variant. On 26 November 2021, the European Centre for Disease Prevention and Control classified the

Omicron B.1.1.529 variant as a variant of concern due to concerns regarding immune escape and potentially increased transmissibility compared to the SARS-CoV-2 delta variant.[8] However, existing studies largely examined COVID-19 during the period when Omicron was not the dominant variant. The risk of venous and arterial thromboembolic events with Omicron COVID-19 is therefore less well studied, particularly in the context of exposure to either prior COVID-19 infection or prior SARS-CoV-2 vaccinations.

7.3 Thromboembolic events and worsening of COVID-19 during the Omicron period

COVID-19 patients with thromboembolic events are at increased risk of worse outcomes, with a systematic review finding a strong association between cardiovascular and thromboembolic events and poor prognosis in COVID-19.[9] However, it has been suggested that the Omicron variant has a milder course and therefore, subsequent coagulopathy risk may differ. This may be further influenced by prior SARs-CoV-2 vaccination. As with measuring the incidence of thromboembolic events themselves, routinely collected data may also be used to describe the risks of worsening in COVID-19 patients during the period when Omicron was the dominant variant.

7.4 Risk factors for thromboembolic events in COVID-19 during the Omicron period

Various patient factors have been associated with worse outcomes in COVID-19 that occurred during the early stages of the pandemic when Omicron was not the dominant variant. Older age, male sex, hypertension, diabetes, and being overweight or obese have all been reported to be associated with an increased risk of hospitalisation and mortality in COVID-19.[10-16] Many of these same factors have also previously been seen to predispose individuals to thromboembolic events.[17,18] In one study a set of pre-existing cardiovascular risk factors were associated with increased mortality in COVID-19, independent of patients´ age and sex.[19] Whilst the associations between such risk factors and thromboembolic events among patients with SARS-CoV-2 variants has been studied, this largely included data on COVID-19 during the early stages of the pandemic when Omicron was not the dominant variant. Information is limited on whether risk factors for venous and arterial thromboembolic events remain the same for COVID-19 associated with the Omicron variant, and to what degree this may be influenced by prior SARs-CoV-2 vaccination and the impact of immunosuppression.

7.5 Contextualising incidence rates for thromboembolic events with COVID-19 during the Omicron period

Two studies have compared incidence rates of thrombosis and thrombocytopenia after vaccination against SARS-CoV-2 and with COVID-19 using data from the United Kingdom and Spain. These studies calculated incidence rates of thromboembolic events in people vaccinated against SARS-CoV-2 and in people with COVID-19 infection and compared them to pre-pandemic rates in a historic background cohort.[20,21] Compared to pre-pandemic rates, standardised incidence ratios were elevated for venous thromboembolism shortly following both initial vaccination against SARS-CoV-2 and for COVID-19 infection, although to a much greater extent with COVID-19 infection. It is uncertain however to what extent evidence generated by these studies is generalisable to COVID-19 infection during the Omicron period.

8 Aim and objectives

The aim of the project is to contextualise the risk of venous and arterial thromboembolic events associated with COVID-19, during the Omicron period and in light of prior SARS-CoV-2 vaccination and infection.

The research objectives which will be addressed incrementally to support the project aim are:

1) To estimate the background incidence rate of venous and arterial thromboembolic events among the general pre-pandemic population.
2) To estimate the incidence rate of venous and arterial thromboembolic events among patients with COVID-19 within 30-, 60-, and 90- and 180-days during the Omicron period, stratified by prior SARS-CoV-2 vaccination and prior infection status.
3) To estimate the incidence rate of venous and arterial thromboembolic events among patients with SARS-CoV-2 vaccination within 30-, 60-, 90- and 180-days, during the Omicron period, stratified by prior infection status.
4) To estimate the association between clinical risk factors and prior SARS-CoV-2 vaccination on the incidence rate of venous and arterial and impact of thromboembolic events on worsening severity of COVID-19 during the Omicron period.
5) To estimate incidence rate ratios for venous and arterial thromboembolic events among patients with COVID-19 and people vaccinated against SARS-CoV-2, compared to the background population using incidence rates estimated in objectives 1 to 3.

9 Research methods

9.1 Study design

Observational cohort study using routinely collected health care data.

9.2 Setting

9.2.1 Data nodes

Data nodes from France, Denmark, Finland, Croatia and selected databases from the DARWIN EU network will be used for the analyses (see section 9.4 Data Sources below for more details). EMA in-house databases in the OMOP common data model may also be used to execute the DARWIN EU study package depending on the final selection and distribution of data sources

9.2.2 Study period

The study period will start depending upon when the Omicron variant became established in Europe commencing from 1st December 2021.[8] The end of the study period will be the last available date of data collection for each contributing dataset.

9.2.3 Study cohorts

Five non-mutually exclusive study cohorts for the analyses will be defined:

1. **General population cohort for pre-pandemic background rates** will have
    - people included in the database as of 1 January 2017 (index date).
    - follow-up for this cohort will run up to 31 December 2019.
2. **Persons tested positive for SARS-CoV-2 or with a clinical diagnosis of COVID-19 during the Omicron period** will
    - have either a positive test result for SARS-CoV-2 or a clinical diagnosis of COVID-19 on or after 1st December 2021 (with the index date being whichever date comes first if both occur)
    - have no positive test result for SARS-CoV-2 or clinical diagnosis of COVID-19 within 3 months prior to the index date.
3. **Persons hospitalised with COVID-19 during the Omicron period** will have
    - a hospitalisation on or after 1st December 2021 (with the index date the date of hospital admission),

- a record of a clinical diagnosis of COVID-19 or a positive test result for SARS-CoV-2 in the period between 3 weeks prior to and up to three days following the index date
- have no COVID-19 hospitalisation within 3 months prior to the index date.

4. **Persons requiring intensive services during a hospitalisation with COVID-19 during the Omicron period** will have
    - intensive services initiated during a hospitalisation as described in cohort 3 (with the index date the date at which intensive services were initiated),
5. **Persons vaccinated against SARS-CoV2 infection**
    - a vaccination record identified by brand (with the index date the date of vaccination record)
    - whether the vaccination record is 1st, 2nd, 3rd or 4th dose.

Cohorts 2 to 5 will be stratified by prior COVID-19 infection status and for cohorts 2 to 4 also by prior SARS-CoV-2 vaccination status. All cohorts will additionally be stratified by whether patients are immunocompromised on the index date. People in each of the five cohorts will be required to have at least a year of observed history in the database prior to their index date. This is to ensure a sufficient time period to identify health conditions and medication use prior to individuals´ index dates.

9.2.4 Follow-up

For cohort 1, follow-up will begin on January 1st 2017 and continue up until the first of: outcome of interest, loss to follow-up, death or December 31st 2019. For cohorts 2 to 5, follow-up will begin on their index date and continue up until the first of: outcome of interest, loss to follow-up, death, or either 30-, 60-, 90- or 180-days after the index date (depending upon the follow-up time period of interest).

9.3 Variables

9.3.1 Exposures

9.3.1.1 Positive test result for SARS-CoV-2

RT-PCR tests have high sensitivity and specificity for SARS-CoV-2. However, as a result of changes in the availability of population wide RT-PCR and home self-reported lateral flow tests during the Omicron period, it may not be possible to exclusively use RT-PCR tests

when identifying positive test results in all datasets.[22] All positive test results for SARS-CoV-2 observable in the database will therefore be included, with documentation of what type of test it was to allow sensitivity analysis restricted to RT-PCR diagnosed patients as needed. The date associated with the test will be used.

9.3.1.2 Clinical diagnosis of COVID-19

Whilst testing for SARS-CoV-2 was commonly performed in many of the countries that will be represented in this study, clinical diagnoses of COVID-19 were also made for many individuals. Diagnostic codes compatible with COVID-19 will therefore also be identified, with the recorded date being used in the analyses.

9.3.1.3 Hospitalisation with COVID-19

Patients hospitalised with COVID-19 will be identified based on having a hospitalisation along with a confirmatory diagnosis or test result of COVID-19 (both as defined above) within a time window from 21 days prior to admission up to three days following their admission. This time window has been chosen to include those who had the diagnosis made prior to their hospitalisation and to allow for a delay in test results or diagnoses to be made or recorded, while excluding individuals with hospital-acquired COVID-19.

9.3.1.4 Intensive care services during a hospitalisation with COVID-19

Patients who received intensive care services during a hospitalisation with COVID-19 will be identified based on having a hospitalisation where they were admitted to the intensive care unit, received mechanical ventilation, tracheostomy, or extracorporeal membrane oxygenation (ECMO). If the date at which the intervention was initiated is observable in the database, this date will be used as the index date. If the date at which the intervention was initiated is not observed (for example, if such interventions are recorded at time of discharge) then the date of ICU admission, or hospital admission if ICU admission is not recorded, will be used as the index date. Patients will have had a confirmatory diagnosis or test result of COVID-19 (both as defined above) within a time window from 21 days prior to their index date up to three days following their index date. As above, this time window was chosen to include those who had the diagnosis made prior to their hospitalisation and allow for a delay in test results or diagnoses to be made or recorded.

9.3.1.5 Vaccination against SARS-CoV-2

COVID-19 vaccine (ATC codes J07BN) exposure will be defined as the date of a vaccination record. Vaccination exposures will be defined by dose (1st, 2nd, 3rd, 4th etc.) and brand.

9.3.2 Study outcomes

9.3.2.1 Venous thromboembolic events

In the primary analysis, venous thromboembolic events will be identified by diagnostic codes for pulmonary embolism or deep vein thrombosis. In a secondary analysis pulmonary embolism and deep vein thrombosis will be assessed separately. We will also assess portal vein thrombosis, splanchnic venous thrombosis (SVT) and cerebral venous sinus thrombosis separately.

9.3.2.2 Arterial thromboembolic events

In the primary analysis, arterial thromboembolic events will be identified by an acute myocardial infarction or acute ischemic stroke. In a secondary analysis acute myocardial infarction and acute ischemic stroke will be assessed separately. We also identify stroke in general, for which we will include both ischemic, haemorrhagic and non-specifically recorded stroke.

9.3.2.3 Cardiovascular events

Instances of heart failure, cardiac arrhythmia, and angina will be identified. In addition, major cardiovascular events (MACE) will be identified by heart failure, acute myocardial infarction, or stroke, or the occurrence of sudden cardiac death. As a sensitivity analysis, we will require that events were identified by hospitalisation admission or discharge records.

9.3.3 Characteristics of study participants

9.3.3.1 Location at index date

For cohort 2 (Persons tested positive for SARS-CoV-2 or with a clinical diagnosis of COVID-19) and cohort 5 (Persons vaccinated against SARS-CoV2 infection) we will identify whether individuals were currently hospitalised on their index date. All people in the cohort 3 (Persons hospitalised with COVID-19) and cohort 4 (Persons requiring intensive services during a hospitalisation with COVID-19) will be hospitalised on their index date.

9.3.3.2 Demographics

Patients´ age at index date and sex will be identified. Age groups will also be identified using the following groupings: <20; 20-44; 45-54; 55-64; 65-74; 75-84; ≥85 years. For those databases where such information is available, individual or area level socioeconomic status and whether they are living in the community or were a nursing home resident will also be identified.

9.3.3.3 Health conditions pre-index date

Individuals´ history of the study outcomes will be identified over three time periods prior to the index date:

1) 30 days prior to one day prior index date,
2) 365 days prior to one day prior index date,
3) all available days observed up to one day prior to index date.

A range of health conditions, including whether a patient was immunocompromised prior to the index date, will be identified using the same time windows. Among these, the following conditions will be identified: antiphospholipid syndrome, asthma, chronic obstructive pulmonary disease, atrial fibrillation, cancer (excluding non-melanoma skin cancers), venous thromboembolism, myocardial infarction, stroke, transient ischaemic attack (TIA), heart failure, diabetes, chronic kidney disease, chronic liver disease, hypertension, rheumatoid arthritis, thrombophilia, inflammatory bowel disease (Crohn's disease or ulcerative colitis), dementia, alcohol or drug substance misuse and obesity.

9.3.3.4 Medications pre-index date

Pre-existing medication use will be identified using two time windows defined as 183 days to one day prior to index date, and 30 days to 1 day prior to index date. Medications of interest will be identified on the basis of Anatomical Therapeutic Chemical (ATC) codes, with use of the following medications identified: COVID-19 medications (ATC code: J05AB18 molnupiravir, J05AE30 nirmatrelvir ritonavir, J06BD05 sotrovimab), non-steroidal anti-inflammatory drugs (ATC group: M01A, with all descendant codes included), Cox2 inhibitors (M01AH), systemic corticosteroids (H02AB and H02BX), antithrombotic and anticoagulant therapies (B01A), lipid modifying agents (C10), agents acting on the renin-angiotensin system (C09), antineoplastic and immunomodulating agents (L), hormonal contraceptives for systemic use (G03A), tamoxifen (L02BA01), and sex hormones and modulators of the genital system (G03).

9.3.3.5 Immunocompromised at the index date

People who are immunocompromised at the index data will be defined by the recording of certain conditions or certain conditions and treatments prior to index date. People will be considered immunocompromised if they have one or more of the following conditions recorded within 365 days prior to index date:

- HIV/AIDS,
- Hematological malignancies
- Solid malignancies
- Other intrinsic immune conditions

People will be defined as being immunocompromised if they are treated with antineoplastic and immunomodulating agents (as defined in 9.3.3.4) between 183 days to one day prior to index date. People will also be defined as being immunocompromised if they are treated with systemic corticosteroids between 183 days to one day prior to index date and have a recording of the following within 365 days prior to index date:

- Organ transplantations
- Rheumatologic/inflammatory conditions

9.3.3.6 Smoking status pre-index date

Individuals´ smoking status (current smoker, ex-smoker, or non-smoker) will be identified when available. All available history for an individual will be used to identify records of their smoking status, with the most recent record included in the analysis.

9.3.3.7 Medications on or post-index date

We will also identify medication use on or after the index date first up to 30-days. For each medication of interest, we will group users into prevalent and new users. The following medications will be identified where available: anticoagulants, anti-platelet drugs, thrombolytic agent, or transfusion with blood products or immunoglobulins, and COVID-19 medications (molnupiravir, nirmatrelvir ritonavir, sotrovimab).

9.4 Data sources

For this study, we will include routinely collected healthcare data from databases. These databases are summarised in Table 1 below. Some of these databases have been mapped to the OMOP CDM, whilst others are in the process, with data also in native format.

Table 1: Data sources accessible for analysis

Data node	Country	Description
Health Data Hub	France	The Health Data Hub has access to the SNDS (Système National des Donnés de Santé), which is the French nationwide healthcare database. It currently covers the overall French population (about 67 million persons) from birth (or immigration) to death (or emigration), even if a subject changes occupation or retires. Using a unique pseudonymised identifier, the SNDS merges all reimbursed outpatient claims from all French health care insurance schemes (SNIIRAM database), hospital-discharge summaries from French public and private hospitals (PMSI database), and the national death register. SNDS data are available since 2006. Only the data on the variables used for the analyses will be taken into consideration on the study population needed to achieve the objectives of the use case. The same variables will be included: • General characteristics: gender, year of birth, area of residence, etc. • Death: month, year, and cause • Long-Term Disease registration associated with an ICD-10 diagnostic codes • Outpatient reimbursed healthcare expenditures with dates and codes (but not the medical indication nor result): visits, medical procedures, nursing acts, physiotherapy, lab tests, dispensed drugs, and medical devices, etc. F • Inpatients details: primary, associated ICD-10 diagnostic codes resulting from hospital discharge summaries with the date and duration of the hospital stay, the performed medical procedures, and the related costs.
Danish Health Data Authority	Denmark	Danish Vaccination Register, hospital-based diagnoses from the Danish National Patient Register, and the unique personal identifier issued by the Danish Civil Registration System allow to create a full population cohort with information on vaccination status of all individuals in Denmark. The unique identifier ensures completeness concerning linkage to registered medical incidents because it is used for registration of all contacts within the Danish healthcare system. Information on SARS-CoV-2 polymerase chain reaction (PCR) test results can be obtained from the Danish Microbiology Database including antigen test data performed by test centers.. The Danish patient register provides information on all inpatient stays and hospital outpatient clinic contacts (including emergency room visits) and provide information on rates of hospital contacts for a range of prespecified cardiovascular and haemostatic diagnoses, including arterial events, venous thromboembolism, thrombocytopenia/coagulation disorders, and bleeding events. Danish registers contain information on: -vaccination status -PCR-test and antigen test results -hospitalisation, ICU admission, medical diagnoses, covariates - prescriptions filled in community pharmacies
Croatian Institute of Public Health	Croatia	Data come from the National Public Health Information System (NAJS), eVaccination database and Integral Central Healthcare Information System of the Republic of Croatia (CEZIH) and is linked via unique personal identifier. CIPH has information on vaccination status, PCR-test results, hospitalisations, medical diagnoses (ICD10), medications (ACT) and covariates for a limited number of cases
Finnish Institute of Health and Welfare	Finland	The national Care Register for Health Care (Hilmo) collects data on the activities of health centres, hospitals and other institutions providing inpatient care and on the clients treated in them as well as on home-nursing clients for the purposes of statistics, research and planning. The unique personal identifier issued by the Population Information System allow to create a full population cohort The Care Register contains whole-population - data including:

		1. patients discharged from inpatient care; 2. count of patients in inpatient care in health centres and hospitals on 31 December; 3. day surgeries; and 4. specialised outpatient care. The Finnish Institute for Health and Welfare (THL) maintains a Finnish national vaccination register. Vaccination data are collected directly from patient record systems. The vaccination register covers vaccinations given in public primary health care. The data is also obtained on vaccinations administered in specialist medical care and private health care. Data on causes of death can be retrieved from Statistics Finland. Demographic data can be retrieved from Statistics Finland and the Digital and Population Data Services Agency. Data on purchases of prescribed medications and reimbursements for medical expenses can be retrieved from the Social Insurance Institution. The different registers can be linked using a personal identity code issued to all Finnish citizens, as well as permanent residents. Most of registers in questions have been established already several decades ago. Only the data on the variables used for the analyses will be extracted on the study population needed to achieve the objectives of the use case. The same variables will be included: • General characteristics: gender, year of birth, area of residence, etc. • Death: month, year, and cause • Long-Term Disease registration associated with an ICD-10 diagnostic codes • Outpatient reimbursed healthcare expenditures with dates and codes (but not the medical indication nor result): visits, medical procedures, nursing acts, physiotherapy, lab tests, dispensed drugs, and medical devices, etc. • Inpatients details: primary, associated ICD-10 diagnostic codes resulting from hospital discharge summaries with the date and duration of the hospital stay, the performed medical procedures, and the related costs.
Darwin EU		Databases TBC

*The EMA in-house databases in the OMOP common data model may also be used to execute the DARWIN EU study package depending on the selection and distribution of data sources.

9.5 Study size

For each database, all individuals that satisfy the eligibility criteria for a study cohort will be included.

9.6 Data management

Native data analysis

In non-DARWIN EU data nodes, data in native format will be examined for completeness and will be structured in a way relevant to each data partner to analyse the chosen objectives. Analyses of native data may generate codes using different software.

OMOP CDM

The DARWIN EU network databases used in this study will have data standardised to the OMOP CDM. The OMOP CDM enables the use of standardised analytics and tools across the network since the structure of the data and the terminology system is harmonised. The

OMOP CDM is developed and maintained by the Observational Health Data Sciences and Informatics (OHDSI) initiative and is described in detail on the wiki page of the CDM: https://ohdsi.github.io/CommonDataModel/ and in The Book of OHDSI: http://book.ohdsi.org

The DARWIN EU analytic package for this study will be written in R. These packages may also be run in OMOP converted data within non-DARWIN EU data nodes, which will allow both within country and between country comparisons.

Each data node will execute the study code against their database containing patient-level data to generate a set of results which will only contain aggregated data. The results from each of the contributing data sites will then be combined in tables and figures for the study report. Prior to study execution on native and OMOP converted data, quality assessment will be undertaken using tools as described in section 9.8.1.

9.7 Data analysis

9.7.1 Descriptive statistics

The observed characteristics of each study population overall, and stratified by age, sex, prior COVID-19 diagnosis, prior SARS-CoV-2 vaccination, whether patients were hospitalised or whether patients were immunocompromised on their index date will be reported. The time at risk observed along with the number of events observed over follow-up will be summarised for each study population overall and by age and sex as well as stratified by whether individuals´ had a prior COVID-19 diagnosis, prior SARS-CoV-2 vaccination, where hospitalised on the index date or where immunocompromised on the index date. The proportion of missing data for a given characteristic, for example relating to smoking status, will also be reported.

9.7.2 Incidence of study outcomes

The incidence of each study outcome described in section 9.3.2 will be estimated during 30-, 60-, and 90- and 180-days following the index date for each cohort of interest with 95% confidence intervals. The 90-day cumulative incidence of study outcomes will be estimated. Given the risk of mortality among patients with COVID-19, particularly among those hospitalised, the competing risk of mortality will be accounted for by estimating cumulative incidence functions. If death is not available, cumulative incidence will be estimated using the Kaplan-Meier approach. As well as estimating the incidence of outcomes for each study cohort as a whole, incidence will also be estimated by age group stratified by sex. Other

stratifications by prior COVID-19 diagnosis, prior SARS-CoV-2 vaccination, whether patients were hospitalised on the index date or where immunocompromised on the index date will also be performed.

9.7.3 Assessing the association between risk factors for thromboembolic events and COVID-19 during the Omicron period

To assess the association between potential risk factors on the incidence of venous and arterial thromboembolic events among patients with COVID-19 during the Omicron period, cause-specific Cox models will be used to calculate hazard ratios for the incidence of venous and arterial thromboembolic events for each of the COVID-19 cohorts. Adjusted models will evaluate potential predictors including age, sex, prior COVID-19 infection status, prior vaccination status, cancer, whether patients were immunocompromised on the index date, prior use of antithrombotics, prior use of corticosteroids, and pre-index comorbidities listed in section 9.3.3.3.

9.7.4 Risks of COVID-19 "worsening" (hospital admission or death) stratified by thromboembolic event occurrence during the Omicron period

A multistate-type modelling approach will be used to assess risks of COVID-19 worsening during the period when Omicron was the dominant variant, stratified by thromboembolic event occurrence. Multistate models allow for a consideration of individuals progression to multiple events of interest, extending on competing risk models by also describing transitions to intermediate events.[23] In the context of COVID-19, hospitalisations during a hospitalisation can be considered as key intermediate events between testing positive for SARS-COV-2 or having a clinical diagnosis of COVID-19 in an outpatient setting on the one end to a COVID-19-related death on the other.

The exact structure of the model used will depend on the characteristics of the data sources available to each data partner (which will be examined in detail as part of the use case described in section 9.8.2). Venous thromboembolism and arterial thromboembolism will be assessed separately, as time-dependent exposures for the following transitions:

(1) from outpatient COVID-19 diagnosis or PCR test positive to hospitalised with COVID-19

(2) from outpatient COVID-19 diagnosis or PCR test positive to death (without a COVID-19 hospitalisation in between)

(3) from being hospitalised with COVID-19 to death

Figure 1 shows an exemplar model for assessing COVID-19 worsening stratified by venous thromboembolic event, where individuals would begin by being identified as having a

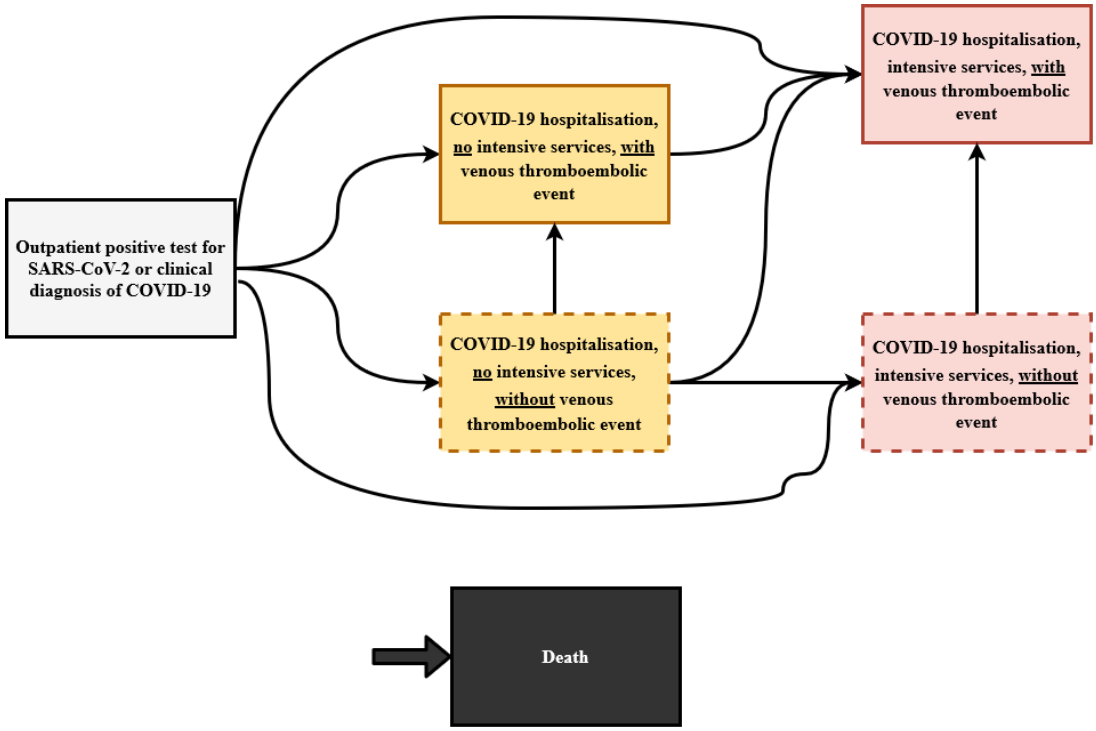
positive test for SARS-CoV-2 or clinical diagnosis of COVID-19 in an outpatient setting and then would progress through the various hospitalisation-related states, and would capture deaths (either after such a hospitalisation or directly after the test, for those individuals who were not hospitalised before their deaths). Cause-specific Cox models within the multistate framework will be used to estimate hazard ratios associated with the risk factors of interest. This approach will allow for the factors of interest to have a different effect by the transition of interest and, where the model includes a state representing deaths, will account for the competing risk of mortality. Models will be adjusted for age and sex.

9.7.5 Contextualising incidence rates for thromboembolic events in COVID-19 during the period when Omicron was the dominant variant

We will calculate crude incidence rate ratios (IRRs) for each event described in section 9.3.2 with 95% confidence intervals, for the vaccinated and COVID-19 cohorts compared against the background general population cohort. This will be done both overall and stratified by key factors including age, sex, prior COVID-19 diagnosis, prior vaccination status and whether patients are immunosuppressed. We will estimate the number of events expected among the vaccinated and COVID-19 cohorts using indirect standardisation by age and sex (with 10-year age bands), using the general population cohort as the standard population. We will calculate standardized incidence rate ratios (SIRs) and 95% confidence intervals comparing observed and expected rates. Rates will be reported overall, and by 10-year age bands, by sex, by prior COVID-19 diagnosis, by prior vaccination status and whether patients were immunosuppressed. A standardized incidence rate ratio above 1 will indicate that the observed rate for a specific outcome is higher than what is expected in the general background population.

Each non-DARWIN EU data node may focus on all or a selected number of objectives in the protocol for conducting native data analysis, according to data availability.

Figure 1. Example multi-state model framework to assess COVID-19 worsening stratified by venous thromboembolic event



9.8 Quality control

9.8.1 General database quality control

Databases in native format

Each of the contributing data node will examine any data provided in their native format for completeness, quality and structure.

Databases in OMOP CDM

For DARWIN EU and other data nodes that will have data in the OMOP CDM, a number of open-source quality control mechanisms for the OMOP CDM have been developed (see Chapter 15 of The Book of OHDSI http://book.ohdsi.org/DataQuality.html). It is expected that data partners with data in the OMOP CDM will run the OHDSI Data Quality Dashboard tool (https://github.com/OHDSI/DataQualityDashboard). This tool provides a number of checks relating to the conformance, completeness and plausibility of the mapped data. Conformance focuses on checks that describe the compliance of the representation of data against internal or external formatting, relational, or computational definitions, completeness in the sense of data quality is solely focused on quantifying missingness, or the absence of data, while plausibility seeks to determine the believability or truthfulness of data values.[23]

Each of these categories has one or more subcategories and are evaluated in two contexts: validation and verification. Validation relates to how well data align with external benchmarks with expectations derived from known true standards, while verification relates to how well data conform to local knowledge, metadata descriptions, and system assumptions. Where possible, the analysis of the same data but in native format will further support validation by allowing the comparison of estimates from the same country and data origin.

9.8.2 Study-specific quality control

Each of the databases with data converted to OMOP will first run an OHDSI cohort diagnostics package (https://github.com/OHDSI/CohortDiagnostics) to identify the exposure cohorts described above. The results will include a summary of codes from the concept sets that are observed in the database and a summary of the concepts that led to entry into particular study cohorts. It also includes a check for ´orphan concepts´, concepts observed in the database that are not included in a concept set of a cohort but perhaps should have been. This will help to inform a consideration of the validity of the exposure cohorts in each of the databases. Subsequently, a cohort diagnostics package will be run to assess the identification of outcomes and patient features in an analogous manner. The analytic study package will then be run against each database. This package will be developed using best practices described in Chapter 17 of The Book of OHDSI (http://book.ohdsi.org/SoftwareValidity.html) and will be made publicly available via GitHub.

9.9 Limitations of the research methods

The study will be informed by routinely collected health care data that is available through existing governance and infrastructure and so data access and quality issues must be considered, and these will be documented to inform the HealthData@EU pilot. For example, cohorts of individuals with COVID-19 will be identified. However, outpatient tested positive cohorts will not capture all those individuals infected, given the variable availability of testing in outpatient settings. The inclusion of a study cohort identified by clinical diagnoses will address some of these limitations.

The included databases will vary in the data elements that they capture. Depending on the dataset, not all study populations will be observed. For example, identifying the intensive services cohort will not be possible where inpatient hospital interventions are not observed.

Similarly, not all outcomes may be available in all databases e.g. deaths. Some datasets may not capture COVID-19 diagnoses or tests occurring outside of hospital. Only the cohorts and outcomes that can reliably be identified will be assessed in the analyses. The data elements that each data partner can provide will be detailed. The contextualisation of incidence rates using a historical background period assumes that coding practices and data recording are similar over time.

9.10 Other aspects

Each data node will be asked to provide data and experience on the process of study approval, data access, data content, data analysis and results consolidation.

10 Protection of human subjects

For this study, participants from various EU member states and the UK will process personal data from individuals which is collected in national/regional electronic health record databases. Due to the sensitive nature of this personal medical data, it is important to be fully aware of ethical and regulatory aspects and to strive to take all reasonable measures to ensure compliance with ethical and regulatory issues on privacy.

All the databases used in this study are already used for epidemiological research and have a well-developed mechanism to ensure that European and local regulations dealing with ethical use of the data and adequate privacy control are adhered to.

In agreement with these regulations, rather than combining person level data and performing only a central analysis with data from multiple member states data in one location, local analyses will be run, which generate non-identifiable aggregate summary results.

Where required, Institutional Review Boards of the respective databases and ethics committees will review the protocol of the study.

Regulatory and ethical compliance

This study was designed and shall be implemented and reported in accordance with the Guidelines for Good Pharmacoepidemiology Practices (GPP) of the International Society for Pharmacoepidemiology, the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines, and with the ethical principles laid down in the Declaration of Helsinki.

This study is fulfilling the criteria of a ‘European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP) study’ and follows the ‘ENCePP Code of Conduct’. The study will be registered in the EUPAS register.

11 Management and reporting of adverse events/adverse reactions

According to the new guideline on good pharmacovigilance practice (EMA/873138/2011 Rev 2*) there is no requirement for expedited reporting of adverse drug reactions from studies with secondary use of data (such as electronic health care databases).

12 Plans for disseminating and communicating study results

Dissemination activities to be undertaken will include mainly, although not exclusively, the creation of a final report, scientific publications, and presentations at conferences.

13 References

1. Bikdeli B, Madhavan M V, Jimenez D, et al. COVID-19 and Thrombotic or Thromboembolic Disease: Implications for Prevention, Antithrombotic Therapy, and Follow-Up: JACC State-of-the-Art Review. J Am Coll Cardiol 2020;75:2950–73. doi:https://doi.org/10.1016/j.jacc.2020.04.031

2. Klok FA, Kruip MJHA, van der Meer NJM, et al. Incidence of thrombotic complications in critically ill ICU patients with COVID-19. Thromb Res 2020;191:145–7. doi:10.1016/j.thromres.2020.04.013

3. Lodigiani C, Iapichino G, Carenzo L, et al. Venous and arterial thromboembolic complications in COVID-19 patients admitted to an academic hospital in Milan, Italy. Thromb Res 2020;191:9–14. doi:10.1016/j.thromres.2020.04.024

4. Llitjos J-F, Leclerc M, Chochois C, et al. High incidence of venous thromboembolic events in anticoagulated severe COVID-19 patients. J Thromb Haemost 2020;18:1743–6. doi:10.1111/jth.14869

5. Katsoularis I, Fonseca-Rodríguez O, Farrington P, Jerndal H, Lundevaller EH, Sund M, Lindmark K, Fors Connolly AM. Risks of deep vein thrombosis, pulmonary embolism, and bleeding after covid-19: nationwide self-controlled cases series and matched cohort study. BMJ. 2022 Apr 6;377:e069590. doi: 10.1136/bmj-2021-069590.

6. Knight R, Walker V, Ip S, Cooper JA, Bolton T, Keene S, Denholm R, Akbari A, Abbasizanjani H, Torabi F, Omigie E, Hollings S, North TL, Toms R, Jiang X, Angelantonio ED, Denaxas S, Thygesen JH, Tomlinson C, Bray B, Smith CJ, Barber M, Khunti K, Davey Smith G, Chaturvedi N, Sudlow C, Whiteley WN, Wood AM, Sterne JAC; CVD-COVID-UK/COVID-IMPACT Consortium and the Longitudinal Health and Wellbeing COVID-19 National Core Study. Association of COVID-19 With Major Arterial and Venous Thrombotic Diseases: A Population-Wide Cohort Study of 48 Million Adults in England and Wales. Circulation. 2022 Sep 20;146(12):892-906. doi: 10.1161/CIRCULATIONAHA.122.060785.

7. Xie J, Prats-Uribe A, Gordillo-Marañón M, Strauss VY, Gill D, Prieto-Alhambra D. Genetic risk and incident venous thromboembolism in middle-aged and older adults following COVID-19 vaccination. J Thromb Haemost. 2022 Dec;20(12):2887-2895. doi: 10.1111/jth.15879.

8. European Centre for Disease Control and Prevention. Communicable Disease Threats report. Week 1, 2-8 January 2022. Available at

https://www.ecdc.europa.eu/sites/default/files/documents/Communicable-disease-threats-report-8-jan-2022.pdf Accessed 14/04/2023

9. Hessami A, Shamshirian A, Heydari K, et al. Cardiovascular diseases burden in COVID-19: Systematic review and meta-analysis. Am J Emerg Med Published Online First: 18 January 2021. doi:10.1016/j.ajem.2020.10.022

10. Burn E, Tebe C, Fernandez-Bertolin S, et al. The natural history of symptomatic COVID-19 in Catalonia, Spain: a multi-state model including 109,367 outpatient diagnoses, 18,019 hospitalisations, and 5,585 COVID-19 deaths among 5,627,520 people. medRxiv 2020;:2020.07.13.20152454. doi:10.1101/2020.07.13.20152454

11. Petrilli CM, Jones SA, Yang J, et al. Factors associated with hospital admission and critical illness among 5279 people with coronavirus disease 2019 in New York City: prospective cohort study. BMJ 2020;369. doi:10.1136/bmj.m1966

12. Reilev M, Kristensen KB, Pottegaard A, et al. Characteristics and predictors of hospitalization and death in the first 11 122 cases with a positive RT-PCR test for SARS-CoV-2 in Denmark: a nationwide cohort. Int J Epidemiol Published Online First: 2020. doi:10.1093/ije/dyaa140

13. Zhou F, Yu T, Du R, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet 2020;395:1054–62. doi:10.1016/S0140-6736(20)30566-3

14. Docherty AB, Harrison EM, Green CA, et al. Features of 20 133 UK patients in hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: prospective observational cohort study. BMJ 2020;369. doi:10.1136/bmj.m1985

15. Gupta S, Hayek SS, Wang W, et al. Factors Associated With Death in Critically Ill Patients With Coronavirus Disease 2019 in the US. JAMA Intern Med Published Online First: 15 July 2020. doi:10.1001/jamainternmed.2020.3596

16. Williamson EJ, Walker AJ, Bhaskaran K, et al. OpenSAFELY: factors associated with COVID-19 death in 17 million patients. Nature Published Online First: 2020. doi:10.1038/s41586-020-2521-4

17. Lowe GDO. Common risk factors for both arterial and venous thrombosis. Br J Haematol 2008;140:488–95. doi:https://doi.org/10.1111/j.1365-2141.2007.06973.x

18. Previtali E, Bucciarelli P, Passamonti SM, et al. Risk factors for venous and arterial thrombosis. Blood Transfus 2011;9:120–38. doi:10.2450/2010.0066-10

19. Collard D, Nurmohamed NS, Kaiser Y, et al. Cardiovascular risk factors are independently associated with COVID-19 mortality: a prospective cohort study. medRxiv 2020;:2020.10.01.20205229. doi:10.1101/2020.10.01.20205229
20. Burn E, Roel E, Pistillo A, Fernández-Bertolín S, Aragón M, Raventós B, Reyes C, Verhamme K, Rijnbeek P, Li X, Strauss VY, Prieto-Alhambra D, Duarte-Salles T. Thrombosis and thrombocytopenia after vaccination against and infection with SARS-CoV-2 in Catalonia, Spain. Nat Commun. 2022 Nov 23;13(1):7169. doi: 10.1038/s41467-022-34669-9.
21. Burn E, Li X, Delmestri A, Jones N, Duarte-Salles T, Reyes C, Martinez-Hernandez E, Marti E, Verhamme KMC, Rijnbeek PR, Strauss VY, Prieto-Alhambra D. Thrombosis and thrombocytopenia after vaccination against and infection with SARS-CoV-2 in the United Kingdom. Nat Commun. 2022 Nov 23;13(1):7167. doi: 10.1038/s41467-022-34668-w.
22. Cheng MP, Papenburg J, Desjardins M, et al. Diagnostic Testing for Severe Acute Respiratory Syndrome–Related Coronavirus 2. Ann Intern Med 2020;172:726–34. doi:10.7326/M20-1301
23. Putter H, Fiocco M, Geskus RB. Tutorial in biostatistics : Competing risks and multi-state models. 2007;:2389–430. doi:10.1002/sim
24. Kahn MG, Callahan TJ, Barnard J, et al. A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data. EGEMS (Washington, DC) 2016;4:1244. doi:10.13063/2327-9214.1244